Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language

James Emil Flege
Department of Biocommunication, University of Alabama at Birmingham, Birmingham, Alabama 35294

(Received 1 January 1990; accepted for publication 23 August 1990)

This study examined whether Spanish–English bilinguals are able to fully differentiate Spanish and English /t/ according to voice-onset time (VOT) if they learn English as a second language (L2) in early childhood. In experiment 1, VOT was measured in Spanish words spoken by Spanish monolinguals, in English words spoken by English monolinguals, and in Spanish and English words spoken by bilinguals who learned English either as young children or as adults. As expected, the Spanish monolinguals produced /t/ with considerably shorter VOT values than the English monolinguals. Also as expected, the late L2 learners produced English /t/ with "compromise" VOT values that were intermediate to the short-lag values observed for Spanish monolinguals and the long-lag values observed for English monolinguals. The early learners' VOT values for English /t/, on the other hand, did not differ from English monolinguals' VOT. The same pattern of results was obtained for stops in utterance-medial position and in absolute utterance-initial position. The results of experiment 1 were replicated in experiment 2, where bilingual subjects were required to produce Spanish and English utterances (sentences, phrases, words) in alternation. The results are interpreted to mean that individuals who learn an L2 in early childhood, but not those who learn an L2 later in life, are able to establish phonetic categories for sounds in the L2 that differ acoustically from corresponding sounds in the native language. It is hypothesized that the late L2 learners produced/t/ with slightly longer VOT values in English than Spanish by applying different realization rules to a single phonetic category.

PACS numbers: 43.70.Ep, 43.70.Kv

INTRODUCTION

The native language (L1) one learns in early childhood and a second language (L2) learned later in life often influence one another. This has been shown to be true for processing in the semantic domain (e.g., Lambert and Rawlings, 1969; Obler and Albert, 1978; Magiste, 1979; Mack, 1986), in the syntactic domain (Blair and Harris, 1981; Mack 1986), and in the phonological domain (Altenberg and Cairns, 1983). It is still uncertain whether such mutual influence holds true in the phonetic domain, at least for individuals who learn two languages in early childhood. The present study explored the degree of independence of L1 and L2 phonetic systems by examining in detail the production of /t/ in Spanish and English by two groups of Spanish–English bilinguals: "early learners" first exposed to English at the age of 5-6 years and "late learners" who began learning English as adults.

The /t/ of Spanish and English differ in two major ways. The Spanish /t/ is formed with the tongue tip and blade against the teeth, whereas the English /t/ is formed with contact against the alveolar ridge (Dalbor, 1980). Voiceless stops in the two languages also differ in glottal–supraglottal timing. The /t/ of Spanish is a voiceless unaspirated stop with short-lag voice-onset time (VOT) values, whereas the /t/ of English is a voiceless aspirated stop with long-lag VOT (Abraham and Lisker, 1973; Williams, 1977a; Flege and Eefting, 1986).

Even though the acoustic differences resulting from these articulatory differences may be detectable (Flege and Hammond, 1982; Flege, 1984; Flege, 1990c), listeners seem to classify realizations of /t/ in Spanish and English as the "same" at a phonological level. For example, Bohn and Flege (1990) found that Spanish monolinguals consistently identified long-lag English [t^h] tokens as /t/ in a two-alternative forced-choice test. English monolinguals identified Spanish short-lag [t] tokens as /t/ in the majority of instances even though they had VOT values that, in an experiment with synthetic stimuli, would be expected to give rise to the perception of /d/ (Williams, 1977b; Flege and Eefting, 1986; see also Forrest and Rockman, 1988). It thus appears likely that Spanish–English bilinguals filter out at least some of the audible acoustic phonetic differences between realizations of Spanish and English /t/ because such differences are not used to contrast meaning in either language (Trubetzkoy, 1939; Weinreich, 1953; Morosan and Jamieson, 1989).

Grosjean (1982, 1985, 1989; Grosjean and Soares, 1986; see also Obler and Albert, 1978) claimed that the two language systems of bilinguals can never operate completely independently of one another because both systems are activated at all times, at least to some degree. This view suggests that it may be impossible for anyone who learns an L2 to prevent pronunciation characteristics of the L1 from influencing their pronunciation of the L2, even if they learned the L2 as young children (see Asher and Garcia, 1969; Thompson, 1984). Flege (1988a, 1990a), on the other hand, hypothesized that complete separation of sounds in the L1 and L2 phonetic inventories is possible, at least for early learners.

According to Flege's speech learning model (henceforth, SLM), a native Spanish-speaking child who is first exposed to English by the age of 5 or 6 years will establish a separate phonetic category for English /t/ in addition to the one already established for Spanish /t/. Spanish speakers who learn English later in life, on the other hand, are not expected to add phonetic categories for "similar" L2 sounds such as English /t/. By hypothesis, the mechanism of equivalence classification prevents them from noting acoustic phonetic differences between Spanish [t] and English [t^h] phones, thereby preventing them from establishing a phonetic category for English /t/. Also by hypothesis, a phonetic category is needed for the accurate production and optimally efficient perception of speech sounds.

Since Spanish /t/ and English /t/ are apt to be identified with one another, an examination of how bilinguals produce these sounds will provide an even stronger test of the "phonetic independence" hypothesis derived from Flege's SLM than would an examination of English sounds that might potentially evade equivalence classification by virtue of differing greatly from any sound in the Spanish phonetic inventory (see Flege, 1987a). Grosjean's "constant dual activation" hypothesis would be supported for the phonetic domain if it were shown that both late *and* early L2 learners were unable to fully differentiate the /t/ sounds in Spanish and English (that is, if both bilingual groups produced English /t/ with significantly shorter VOT values than a group of English monolinguals). Support for a phonetic independence hypothesis (Flege, 1988a, 1990a) would be provided by the finding that late L2 learners but not early L2 learners differed significantly from English monolinguals in producing English /t/.

L2 speech production research has shown that few late learners fully differentiate /p,t,k/ in their two languages if voiceless stops in the L1 are realized with short-lag VOT values and voiceless stops in the L2 are realized with *long*-lag VOT values. Previous studies have shown that many adult L2 learners produce English /p,t,k/ with significantly shorter VOT values than English monolinguals, but with significantly longer VOT values than monolingual native speakers of the learners' L1 (e.g., Flege and Port, 1981; Port and Mitleb, 1980, 1983; Nathan, 1987; Flege, 1987a; Major, 1987. Lowie, 1988). When late learners' VOT values for English /p,t,k/ are intermediate to the values observed for monolingual speakers of the L1 and L2 they are said to have been produced with "compromise" values (Williams, 1980). The seeming limitation on how accurately VOT in English /p,t,k/ is produced also seems to apply to adolescents and older children (Williams, 1979 1980. Suomi, 1980; Flege and Eefting, 1987b; Schmidt, 1988).[1] Flege and Hillenbrand (1984) hypothesized that an upper limit exists on the extent to which late L2 learners can approximate the phonetic norm of English for /p,t,k/ based on the observation that compromise VOT values are common for late L2 learners.

Not *all* late learners produce English /p,t,k/ with compromise values, however. Some have produced English /p,t,k/ with short-lag VOT values resembling those typical for /p,t,k/ in the L1, suggesting they simply produced English words with L1 sounds. It is possible that such individuals fail to detect VOT differences between voiceless L1 and L2 stops. Contrary to the "upper limit" hypothesis of Flege and Hillenbrand (1984), a few subjects in previous studies have been observed to produce English /p,t,k/ with VOT values that equaled or even overshot[2] values for native speakers of English (e.g., Suomi, 1980; Major, 1987; Flege and Eefting, 1987b).

The majority of subjects in previous studies who have produced /p,t,k/ with longer VOT values in English than in their L1 must surely have noted at least some of the acoustic differences distinguishing L1 and L2 stops. It is uncertain at present, however, whether they differed from native speakers of English because their perceptual knowledge of English /p,t,k/ was inaccurate, their ability to *reproduce* what they heard was imperfect, or some combination of both. It is also uncertain from previous research whether early learners are better able than late learners to produce English /p,t,k/ with VOT values resembling those of native speakers.

The results of several studies suggest that early learners may fully differentiate /p,t,k/ in L1 and L2. Williams (1977b) reported that adults who learned both English and Spanish by the age of six years did not differ from English monolinguals in producing English stops, nor differ from Spanish monolinguals for Spanish stops. Mack (1989) found that adults who had learned both French and English by the age of 7 years did not differ from English monolinguals in producing English /t/. Fokes *et al.* (1985) examined English stops spoken by 12 native Arabic children ranging from 2–11 years of age. All but one seemed to have produced English /p/ and /t/ with VOT values that were as long (or longer) than those typical for native English children.

Other research indirectly supports the view that early learners may fully differentiate the /p,t,k/ of their L1 and L2. Native Chinese subjects who began learning English at an average age of 7.6 years were found to produce English sentences with a detectable foreign accent, whereas Spanish subjects who began learning English by the age of 5 to 6 years produced the same sentences without an accent (Flege, 1988b, 1990b). Strength of foreign accent in sentences is known to be inversely related to VOT in English /p,t,k/ (Flege and Eefting, 1987b; Major, 1987). Thus individuals who begin learning English in early childhood, but not those who begin learning English in later childhood or as adults, may produce English /p,t,k/ with authentic VOT values.

The results of other L2 production studies, on the other hand, suggest that even early learners may fail to produce English /p,t,k/ authentically. Caramazza *et al.* (1973) found that native French speakers who began learning English by the age of 7 years produced English /p,t,k/ with significantly shorter VOT values than native speakers of English. Flege and Eefting (1987a) also found that native Spanish adults and children who began learning English L2 by the age of 5 to 6 years produced English /p,t,k/ with significantly shorter VOT values than age-matched groups of native English subjects. These studies suggest that early learners may be unable to fully differentiate /p,t,k/ in L1 and L2, and thus support the view that both the L1 and L2 phonetic systems remain activated to some degree.

There is reason to think, however, that the two studies just cited do not indicate accurately how well early learners may produce L2 stops. It is not certain what kind of English-language input subjects in the Caramazza *et al.* (1973) study received, nor how well they spoke English. Since French was the dominant language spoken in the city where the study was carried out (Montreal), the possibility exists that the early learners examined by Caramazza *et al.* did not receive sufficient native speaker input to enable them to produce English /p,t,k/ authentically. VOT values in English stops spoken with a French accent are typically shorter than VOT values in English stops spoken by English monolinguals (Flege and Hillenbrand, 1984; Flege, 1987a). The French-English bilinguals examined by Caramazza *et al.* may simply have produced English stops with VOT values resembling those in the English stops they had heard. The same might also be true for the Puerto Rican subjects examined by Flege and Eefting (1987a), who were living in a predominantly Spanish-speaking community (Mayaguez, Puerto Rico) at the time of the study.

In summary, previous research has established that late learners are apt to produce English /p,t,k/ with VOT values that are too short for English. It remains uncertain as to whether early learners will also differ from native speakers of English, or if they will fully differentiate corresponding L1 and L2 stops. Few previous studies have examined whether learning an L2 affects how bilinguals produce stops in their L1. It appears that no previous study has directly compared the production of L2 stops by early and late learners. Thus the purpose of the present study was to determine how closely early and late learners would resemble native speakers of English in producing English /t/, and whether learning English would affect their production of Spanish /t/.

Experiment 1 examined VOT values in utterance-initial and utterance-medial stops in Spanish and English words that were read from lists. Experiment 2 replicated and extended experiment 1. Its aim was to determine if the same Spanish versus English VOT differences seen in experiment 1 would be observed when subjects were required to produce Spanish and English utterances in alternation. In Sec. III, possible underlying bases for the differences between early and late learners observed in experiments 1 and 2 are discussed in the context of Flege's (1988a, 1990a) speech learning model.

I. EXPERIMENT 1

A. Methods

1. Subjects

Two groups of monolinguals and two groups of bilinguals (six males and four females per group) participated as paid subjects. As summarized in Table I, the Spanish and English monolinguals differed little in mean age (30 vs 26 years). The English monolinguals were students at the University of Texas. The Spanish monolinguals were recruited at a refugee center in Austin, TX. Most had lived in the U.S. for less than 3 months at the time of testing. Only a few of them reported having studied English in school.

Subjects in the other two groups were native speakers of Spanish who learned English as an L2. The early learners indicated that they were first exposed massively to English when they started school in Texas at the age of 5–6 years. Four of the early learners were born in Mexico, the rest in Texas border towns. Only individuals who had native English teachers in the first three primary grades, and/or had a majority of native English classmates in those grades, were included in the early L2 group. The early learners reported being unable to speak English when they began school, a claim that seems reasonable in light of demographic data (see Fernandez and Molinet-Molina, 1988). A study by Flege (1990b) showed that, even though English was their second language, the early learners spoke it without accent. The Spanish-speaking research assistant who recorded the early learners in Austin indicated that their Spanish was also unaccented, but this was not tested formally.

TABLE I. Characteristics of monolingual native speakers of Spanish and English who participated in experiment 1. "POB" indicates place of birth.

Subject	Age	Sex	Spanish POB	Age	Sex	English POB
1	40	M	Mexico City	41	F	Bethesda, MD
2	23	F	Monterey	22	F	Tampa, FL
3	46	M	Veracruz	20	F	Ft. Benning, GA
4	26	M	En Escinapa	21	M	Houston, TX
5	48	M	Mexico City	22	M	Farmers Branch, TX
6	17	M	San Luis Potosi	44	M	Port Arthur, TX
7	17	M	Mexico City	20	M	Akron, OH
8	31	F	Frontera	21	M	New London, CT
9	20	F	San Luis Potosi	26	F	Hondo, TX
10	32	F	Frontera	21	M	Cleburn, TX
M	30			26		

The late learners did not begin learning English until they were adults. Four of them were recorded in Austin, the remaining six in Birmingham. The late learners differed from the early learners principally according to the *age* of L2 learning. However, the two groups of bilinguals differed in other ways, as summarized in Table II. Compared to the early learners, the late learners were somewhat older (34 vs 29 years), had less formal education in English (6 vs 13 years), and spoke English somewhat less on a daily basis according to self-report (66% vs 82%).[3] The late learners had arrived in the U.S. at a much later average age (20 vs 2 years), and so had lived there for a shorter total period of time than the early learners (14 vs 21 years).

2. Materials and procedures

Owing to phonological differences between Spanish and English, it was not possible to find lists of matched English and Spanish words. The words chosen, however, were all disyllabic and had vowels of approximately the same quality following the word-initial /t/. The native Spanish subjects read Spanish words at the end of the Spanish carrier phrase "Tengo un_." Half of the Spanish words were followed by /i/ (*tigre, tipo, tiro, timbre*), the other four by /ɛ/ (*tema, termo, templo, texto*).[4] The English monolinguals and the Spanish–English bilinguals read English words with /i/ or /ɛ/ (*t-bone, teapot, teabag, t-shirt, teller, temple, textbook, tempo*) at the end of the English carrier phrase "Take a_."

The Spanish and English materials were elicited in the

TABLE II. Characteristics of the early and late L2 learners who participated in experiment 1. The 14 subjects who later participated in experiment 2 are marked by an asterisk.

Subject	Age	Sex	POB[a]	EDU[b]	AOA[c]	LOR[d]	PER[e]
			Early L2 Learners				
1*	23	F	Edinburgh, TX	12	0	23	90
2*	26	M	Mexico City	12	6	20	70
3*	19	M	Guadalajara	13	5	14	80
4*	20	F	Farr, TX	13	0	20	80
5	23	F	Dallas, TX	12	0	23	90
6*	20	F	Obregon	11	3	15	80
7*	23	M	unknown	13	1	22	75
8	21	F	Orio	13	0	21	90
9	24	M	Laredo, TX	16	0	24	80
10*	26	M	Taft, TX	13	0	26	85
M	29			13	2	21	82
			Late L2 Learners				
Subject	Age	Sex	POB[a]	EDU[b]	AOA[c]	LOR[d]	PER[e]
1*	28	M	Mexico City, Mex.	0	18	10	70
2*	23	M	Juarez, Mex.	6	13	10	50
3*	37	F	Monterrey, Mex.	10	12	25	50
4*	19	F	Matamoros, Mex.	7	11	8	70
5*	41	M	Chile	6	24	17	73
6	38	F	Ecquador	12	14	24	90
7	53	F	Panama	0	35	18	95
8	28	M	Nicaragua	7	18	10	50
9*	33	M	Guatemala	5	26	7	75
10*	40	F	Chile	4	26	14	35
M	34			6	20	14	66

[a] Place of birth.
[b] Years of formal instruction in English.
[c] Age of arrival in the U.S., in years.
[d] Length of residence in the U.S., in years.
[e] Self-estimated percentage daily use of English.

appropriate languages by bilingual research assistants.[5] The monolingual subjects produced only the Spanish or English materials, whereas the bilinguals produced both in counterbalanced order. Instructions were given to the monolinguals in Spanish or English, as appropriate. The bilingual subjects assigned to produce the Spanish materials first heard the Spanish instructions, and vice versa. The subjects were told that the experiment examined speech, but not that their production of /t/ would be assessed. They were instructed to read each sentence "as if talking to...friends" at a constant speaking rate and loudness level, and to repeat any utterance with which they were unsatisfied. The subjects said the number of each utterance (in the appropriate language), paused, then produced the utterance. The reading task was modeled at a moderate speaking rate on the instruction tape using a list of utterances resembling those on the randomized lists.

3. Measurements

Each of the test words occurred three times on the Spanish and English lists. A total of 14 utterances from the middle of each list were digitized at 10 kHz. Each utterance contained two word-initial /t/ tokens, one in absolute utterance-initial position (i.e., the /t/ in *take* and *tengo* in the carrier phrases) and the other in utterance-medial position (in the /t/ initiating test words at the end of the carrier phrases). VOT was measured to the nearest 0.1 ms from the screen of a graphics terminal from the beginning of the release burst to the first positive peak in the periodic portion of the waveform.

A total of four mean VOT values were calculated for each subject. Mean values for /t/ in the test words beginning with /i/ and /ε/ were calculated, most based on seven observations. The few missing data points were due to the absence of a visible release burst, which made it impossible to measure VOT. Two mean values were calculated for stops in utterance-initial position. One was for utterances ending in test words with /i/, the other for utterances with /ε/ test words.

Measurement reliability was assessed using the test–retest approach. The assistant who measured VOT remeasured 20 randomly selected utterance-initial and 20 utterance-medial /t/ tokens several weeks later. The average (unsigned) difference between the two sets of measurements was 1.5 mm for both the utterance-initial and utterance-medial stops. The largest difference noted was only 5.3 ms so, of course, the first and second sets of measurements were highly correlated ($r = 0.997$ for utterance-initial stops, $r = 0.999$ for utterance-medial stops).

B. Results

1. Utterance medial stops

Figure 1 (bottom) shows the mean VOT values for Spanish and English /t/ tokens that were produced in utterance-medial position. The values shown here have been averaged across the /i/ and /ε/ contexts, the effect of which will be discussed below. As expected, the monolingual English speakers' /t/ had substantially longer VOT values than that of the Spanish monolinguals (64 vs 22 ms). Also as expected, both the early and the late learners produced /t/ with longer VOT values in English than Spanish. The late learners produced English /t/ with shorter VOT values (40 ms) than the English monolinguals, whereas the early learners' English /t/ had the same mean values (viz. 64 ms) as the English monolinguals'. The mean VOT values obtained for each of the 30 subjects who produced the English speech material were submitted to a (3) group × (2) vowel context ANOVA, which yielded a significant group main effect [$F(2,27) = 14.0$, $p < 0.05$]. Newman–Keuls *post hoc* tests revealed that the English monolinguals and early learners produced English /t/ with significantly longer VOT values than the late learners, but did not differ significantly from one another ($p < 0.05$).

The early and late learners did not differ in producing stops in their L1. The subjects in both bilingual groups produced Spanish /t/ with an average VOT of 20 ms, which was slightly shorter than the mean value observed for Spanish monolinguals (viz. 22 ms). The mean VOT values obtained for the 30 subjects who produced Spanish words in the context of /i/ and /ε/ were submitted to a (3) group × (2) vowel context ANOVA. The group factor was nonsignificant [$F(2,27) = 0.447$]. This suggests that having learned Eng-

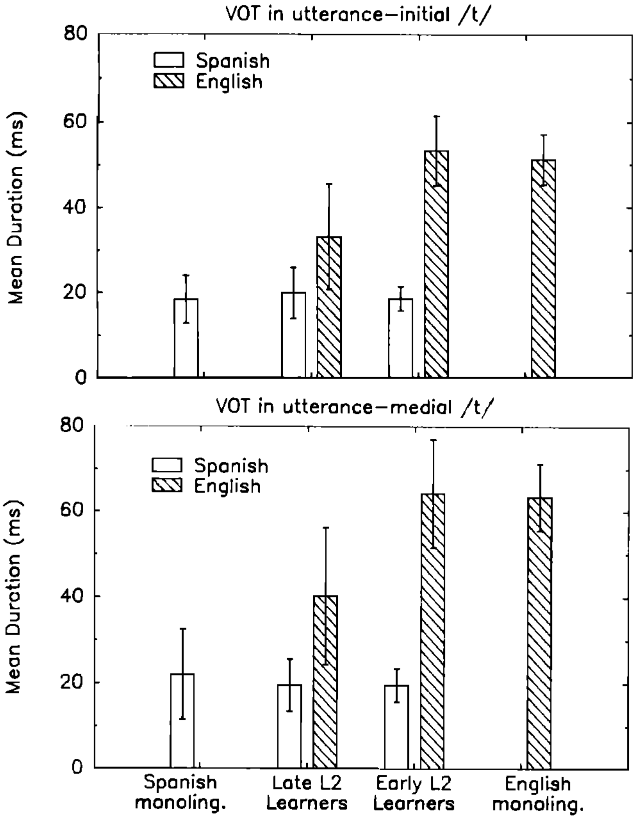

FIG. 1. (top) VOT in the stops initiating Spanish and English carrier phrases. (bottom) The mean VOT value for the word-initial stops in Spanish and English test words beginning in /t/ that were spoken at the end of carrier phrases, in ms. Each mean VOT value is based on ten individual subject means which, in turn, were based on up to seven observations in each of two vowel contexts (/i/,/ε/). The error bars enclose + / − one standard deviation.

lish did not influence how the bilinguals produced stops in their L1.

A (2) group × (2) language ANOVA was carried out to determine if the two groups of bilinguals differed significantly in terms of how well they differentiated the /t/ of Spanish and English. The analysis yielded a significant interaction [$F(1,18) = 12.9, p < 0.05$], which was followed up by tests of the simple main effect of group. These tests showed that both the early and the late learners produced /t/ with significantly longer VOT values in English than Spanish [$F(1,9) = 138.1, 14.3; p < 0.05$]. The interaction was probably due, therefore, to the fact that the early learners produced a substantially larger VOT difference between Spanish and English /t/ than the late learners (44 vs 20 ms).

2. Utterance-initial stops

Figure 1 (top) shows the mean VOT values of Spanish and English /t/ tokens spoken in absolute utterance-initial position. The monolingual English speakers produced /t/ with substantially longer VOT values than the Spanish monolinguals (51 vs 18 ms). The early learners produced English /t/ with about the same mean VOT value (53 ms) as the English monolinguals, whereas the late learners produced English /t/ with a shorter mean value (33 ms). The group factor was significant in the ANOVA examining the English /t/ tokens [$F(2,27) = 15.2, p < 0.05$]. A *post hoc* test revealed that the English monolinguals and the early learners had significantly longer VOT values than the late learners but did not differ from one another ($p < 0.05$).

The early and late learners produced Spanish /t/ with VOT values that were about the same on the average as the Spanish monolinguals' mean values (19, 20 vs 18 ms). Not surprisingly, the group factor was nonsignificant in the ANOVA examining Spanish stops [$F(2,27) = 0.28$].

A (3) group × (2) language ANOVA was carried out to examine the bilingual subjects' production of absolute utterance-initial stops in Spanish and English. This analysis yielded a significant interaction [$F(1,18) = 18.2, p < 0.05$]. As in the analysis of utterance-medial stops, both the early and the late learners produced /t/ with significantly longer VOT values in English than Spanish [$F(1,9) = 225.3$, 8.75; $p < 0.05$]. The interaction was probably due to the fact that the early learners produced a larger Spanish versus English VOT difference than the late learners (34 vs 13 ms).

3. VOT variability

To determine if the bilingual subjects were more variable in producing English /t/ than the English monolinguals, intersubject variability was examined. Figure 2 shows the mean VOT values obtained for individual subjects in utterance-medial and utterance-initial English /t/ tokens in the two vowel contexts. Visual inspection of this figure suggests that intersubject variability may have been greater among the nonnative than native subjects. However, an F_{max} test showed that the assumption of homogeneity of variance for subjects in the three groups could not be rejected either for utterance-medial stops ($F_{max} = 1.63$ with three variances and nine *df* per variance) or for stops in absolute utterance-initial position ($F_{max} = 2.38$)

Most of the early learners produced English /t/ in utterance-medial position with mean VOT values that fell within or exceeded the range of values observed for native speakers of English (viz. 51–76 ms), whereas most of the late learners had mean VOT values that were less than the smallest value observed for any of the English monolinguals. Each of the early learners produced mean values for /t/ in absolute utterance-initial position that fell within or exceeded the English range (viz. 34–63 ms) whereas only about half of the late learners' mean values fell within the English range. Some late learners produced English /t/ in both utterance-medial and utterance-initial position with mean VOT values of about 20 ms, suggesting that they used a Spanish /t/ in English. All of the early learner's mean values exceeded 30 ms and could therefore be termed "long-lag" stops.

A second method was used to determine if the bilingual subjects were more variable in producing English /t/ than the English monolinguals. The standard deviations (s.d.) associated with each of the four mean VOT values obtained for each subject was calculated. Overall, the native English speakers' s.d. values differed little from those of the early and late learners (8.9 vs 9.8, 8.3). The s.d. values associated with the production of English /t/ were submitted to a (3) group × (2) vowel context × (2) utterance position ANOVA, which yielded a significant three-way interaction were nonsignificant in the analyses of both Spanish /t/ and English /t/.

The vowel context effects probably had an aerodynamic basis. It appears that the time needed for the transglottal pressure drop needed for spontaneous voicing may take longer to reestablish itself after the release of stops that precede vowels formed with tight lingual–palatal constriction than vowels without such a constriction.[7] Vowel context effects similar to the ones reported here for Spanish and English stops have been noted in previous studies of stop production by English monolinguals (Port and Rotunno, 1979; Weismer, 1979). The vowel context effect will not be discussed further since it does not appear to have resulted from language-specific phonetic learning.

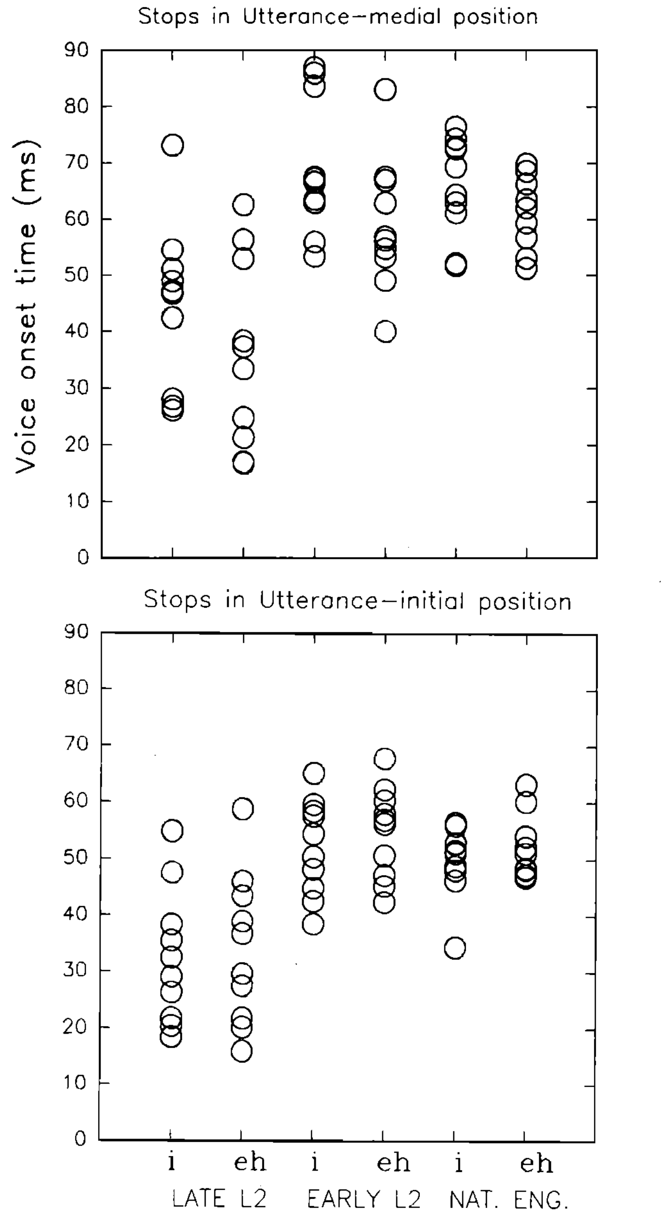

FIG. 2. (top) The mean VOT values obtained for /t/ as produced in the word-initial position of test words with /i/ and /ɛ/ ("eh") found at the end of a carrier phrase, in ms; (bottom) the means for stops in the absolute initial position of the carrier phrase "Take a _ for the utterances with /i/ and /ɛ/ ("eh") test words.

$[F(2,27) = 3.67, p < 0.05]$. The interaction was explored by examining the simple main effect of group at all four vowel × utterance position combinations. The group factor was nonsignificant in every instance ($p > 0.05$), suggesting that the nonnative speakers were no more variable in producing English /t/ than the native speakers.[6]

4. Effect of vowel context

For word-initial English stops in utterance-medial position (that is, in the test words at the end of the carrier phrase), VOT was significantly longer in the context of /i/ than /ɛ/ (60 vs 52 ms) $[F(1,27) = 26.8, p < 0.05]$. For Spanish stops in the same position, VOT was also significantly longer in the context of /i/ than /ɛ/ (23 vs 18 ms) $[F(1,27) = 41.4, p < 0.05]$. The group × vowel interactions

C. Discussion

Spanish /t/ was found to have VOT values of 22 ms in utterance-initial position, and 18 ms in absolute utterance-initial position. These values are slightly longer than values reported by Lisker and Abramson (1964) and Williams (1977a) for Spanish. English /t/ was found to have VOT values of 64 ms in utterance-initial position, and 51 ms in absolute utterance-initial position. These values are also somewhat longer than values reported in some previous studies. For example, Lisker and Abramson (1967) reported a mean value of 48 ms for word-initial /t/ tokens in utterance-medial position, and 45 ms for /t/ tokens initiating words spoken in isolation (and thus in absolute initial position). These differences between studies can probably be attributed to differences in vowel context, degree of stress, and speaking rates.

The primary purpose of experiment 1 was to determine if Spanish–English bilinguals would be able to fully differentiate /t/ in their two languages, that is, to produce a VOT difference equal to the difference between Spanish and English monolinguals. The results suggested that at least some bilinguals are able to fully differentiate their two languages at a phonetic level. Early learners who learned English as young children produced Spanish /t/ with mean VOT values that did not differ significantly from those of Spanish monolinguals, and they produced English /t/ with mean VOT values that did not differ significantly from those of English monolinguals. The same pattern of differences between groups was obtained for stops produced in utterance-medial position and in absolute utterance-initial position.

Late learners who were first massively exposed to English as adults, on the other hand, only partially differentiated English and Spanish /t/. These subjects produced Spanish /t/ with values much like those of Spanish monolinguals. Although they produced English /t/ with significantly longer VOT values than were observed for the Spanish monolinguals' /t/ tokens, they produced English /t/ with significantly shorter VOT values than English monolinguals. Previous studies of L2 production of late learners have also shown such compromise VOT values (e.g., Nathan, 1987; Major, 1987). The finding that few late learners' VOT values for English /t/ closely resembled those of the native speakers despite exposure to native-produced English stops over many years agrees with the observation that adults' L2 pronunciation tends to "fossilize" (Selinker, 1972; Scovel, 1988). It is also consistent with the hypothesis (Flege and Hillenbrand, 1984) that an upper limit exists on how closely late learners can approximate the phonetic norms of an L2 for sounds in the L2 that differ acoustically from corresponding sounds in the L1.

The results obtained here for early learners support an inference drawn from an imitation study by Flege and Eefting (1988). In that study, native speakers of English and early L2 learners produced stops with short-lag and long-lag VOT values when imitating the short-lag and long-lag members of a synthetic VOT continuum ranging from /da/ to /ta/. Late learners, on the other hand, seldom produced long-lag VOT values when imitating stimuli from the long-lag end of the continuum. This finding suggested that the early L2 learners had established a phonetic category for the long-lag /t/ realizations of English even though they produced English /p,t,k/ with compromise VOT values when speaking spontaneously. The authors concluded that the early learners may have produced English /p,t,k/ with "accented" VOT values because they had received foreign-accented input as young children.

The results presented here must be interpreted cautiously for a number of reasons. First, the study did not examine conversational speech. Second, only eight English words were examined, and these were not evaluated for degree of subjective familiarity. If the effects of phonetic learning diffuse gradually across the lexicon, as for children learning English as an L1 (Ferguson, 1986), then the accuracy of VOT in an English stop spoken by L2 learners may depend on when the word containing that stop was first encountered.

Another reason for caution is that the two bilingual groups compared in experiment 1 differed according to factors in addition to age of L2 learning. The early learners were likely to have received much more native-speaker phonetic input than the late learners (see Table II). It is unlikely, however, that a lack of phonetic input *per se* could account for why the late learners produced English /t/ with significantly shorter VOT values than the English monolinguals. Each of them had lived in the U.S. for at least 7 years, and all but one of them reported using English at least half of the time on a daily basis.

It is also unlikely that differences in amount of L2 input could account for the observed VOT differences between the two bilingual groups. Previous research has shown that age of learning is the single most important determinant of how well an L2 will be pronounced. Length of residence, which is often used as a gross estimate of amount L2 phonetic input, has often been shown to exert a much smaller, and usually nonsignificant, effect on global foreign accent (Tahta *et al.*, 1981; Seliger *et al.*, 1975; Oyama, 1976; Purcell and Suter, 1980; Thompson, 1984; Flege, 1990b). Global foreign accent, in turn, is known to be significantly correlated with VOT in stops spoken by nonnative speakers (Flege and Eefting, 1987a; Major, 1987). A number of studies have shown little or no difference for VOT in stops spoken by L2 learners who had just arrived in the U.S. or had resided there for 3 or more years (Williams, 1979, 1980; Flege and Port, 1981; Port and Mitleb, 1983).

A study of intrasubject variability showed that neither the early nor the late learners had greater token-to-token variability in producing /t/ than the English monolingual subjects. There was somewhat more intrasubject variability among the late learners than among the early learners and English monolinguals, but the differences were nonsignificant. An inspection of individual subject data suggested that the mean value reported for the late learners' production of English /t/ did not adequately represent *all* subjects in that group. Although most subjects produced English /t/ with the expected compromise values, some seemed to have produced English words with a "Spanish" /t/ (i.e., with short-lag VOT values of about 20 ms) and a few managed to produce English /t/ with long-lag VOT that fell within the range of values observed for the native English subjects. It is uncertain whether these individual differences were due to differences in underlying phonetic organization, or in the modulation of phonetic parameters.

The results of previous research with late learners reviewed in the Introduction suggested the possibility that the ability to accurately produce voiceless aspirated stops in an L2 may be normally distributed. This observation must be considered tentative for several reasons, however. First, as alluded to above, most previous studies have examined speech that was read rather than spoken conversationally. Such an elicitation procedure is likely to increase the likelihood that articulation strategies may obscure normal patterns of production. Second, relatively few previous studies have reported individual subject data. Most L2 production studies have simply presented mean values for groups of subjects.

II. EXPERIMENT 2

The purpose of this experiment was to replicate and extend experiment 1. In experiment 1, Spanish–English bilingual subjects read lists of English and Spanish utterances in counterbalanced order. The research assistants who elicited the data switched between the two languages, as appropriate, to reduce the artificiality of the speaking situation. However, bilinguals seldom confine themselves to speaking L1 and L2 in distinct, nonoverlapping blocks. They typically switch between their two languages, at least with interlocutors familiar with both languages, as were the assistants who elicited the data (Grosjean, 1982). Therefore, in the present experiment, some of the bilingual subjects from the first experiment produced Spanish and English in alternation.

A. Methods

1. Subjects

Of the bilingual subjects in experiment 1, all but two early learners participated in this experiment. Data for a third early learner were not usable owing to an equipment malfunction. To ensure an even number of subjects in the two bilingual groups, three late learners were eliminated by random selection. The seven early and seven late learners for whom data will be reported are indicated in Table II. This subset of subjects differed in much the same way described earlier in terms of chronological age, formal education in English, self-estimated daily usage of English and—most importantly—age of L2 learning. The early learners in experiment 2 began learning English at the age of 5 to 6 years, whereas the late learners began learning English when they arrived in the U.S. at an average age of 19 years.

2. Speech materials

VOT was measured in the same English and Spanish words that were examined in experiment 1. The words were produced in three consecutive conditions designated the "sentence," "phrase," and "word" conditions. In the sentence condition, disyllabic test words were produced at the end of an English or Spanish carrier phrase, as appropriate ("Take another word such as_"; "Tengo palabras como_"). In the phrase condition, the test same words were produced at the end of an English or Spanish phrase ("Take a_";"Tengo un_"). In the third condition, isolated Spanish and English test words were produced in alternation. The same random order (that of experiment 1) was used in all three conditions.

3. Procedures

The experiment was carried out by the same bilingual research assistants as in experiment 1. Half of the subjects heard recorded instructions in English, half in Spanish. The subjects were told that they would produce English and Spanish sentences, phrases, and isolated words in alternation. To distinguish between languages, the Spanish materials were highlighted on the written lists used to elicit production. In the isolated word condition, language identity was redundantly specified by placing the letters "E" or "S" in front of each word.

The onset of each utterance in the three conditions was regulated by a light-flashing device. The subjects were instructed to time the onset of successive utterances to coincide with the light flashes. The interval between flashes was 3.2 s in the sentence condition, 1.8 s in the phrase condition, and 1.1 s in the word condition. Pilot tests revealed that these were the shortest intervals that some subjects could accommodate. The subjects were permitted to practice with the sentence material before data collection began. The subjects were told to skip an utterance altogether if they lost the rhythm they had established. In several instances the subjects came to a complete halt. The conditions in which this happened were rerun.

4. Measurement

The VOT of word-initial /t/'s in Spanish and English test words were measured to the nearest 0.1 ms from oscillographic displays. A total of 20 word-initial /t/ tokens from the middle of the lists were measured for each subject in each condition. From these, four mean VOT values per condition were calculated: two for English /t/ tokens (one for words with /i/, one for words with /ɛ/) and two for Spanish /t/ tokens (one for words with /i/, one for words with /ɛ/). Most of the mean values were based on five observations; none was based on fewer than three tokens. The assistant listened for pauses between utterances before making the VOT measurements. The /t/'s were not measured in the 1.6% of sentence-condition utterances judged to have been preceded by a pause; the small percentage of utterances in the phrase and word conditions preceded by a perceived pause (3.1% in both instances) were also excluded from analysis.

B. Results

1. VOT

The mean VOT values obtained for the two bilingual groups are shown in Fig. 3. Averaged across the three switching conditions and two vowel contexts, the early learners produced a substantially greater difference between English /t/ and Spanish /t/ (60 vs 21 ms) than the late learners (38 vs 22 ms). The mean VOT values were submitted to a (2) group×(3) switching condition×(2) language×(2) vowel ANOVA with repeated measures on the last three factors. This analysis yielded a significant group ×language interaction [$F(1,12) = 12.0$, $p < 0.05$], which was explored by tests of simple main effects. Both the early and the late learners were found to produce /t/ with significantly longer VOT values in English than Spanish [$F(1,6) = 163.1, 8.72$, $p < 0.05$]. The source of the interaction was therefore likely to have been the greater magnitude of the Spanish versus English difference for the early than the late learners (41 vs 16 ms).

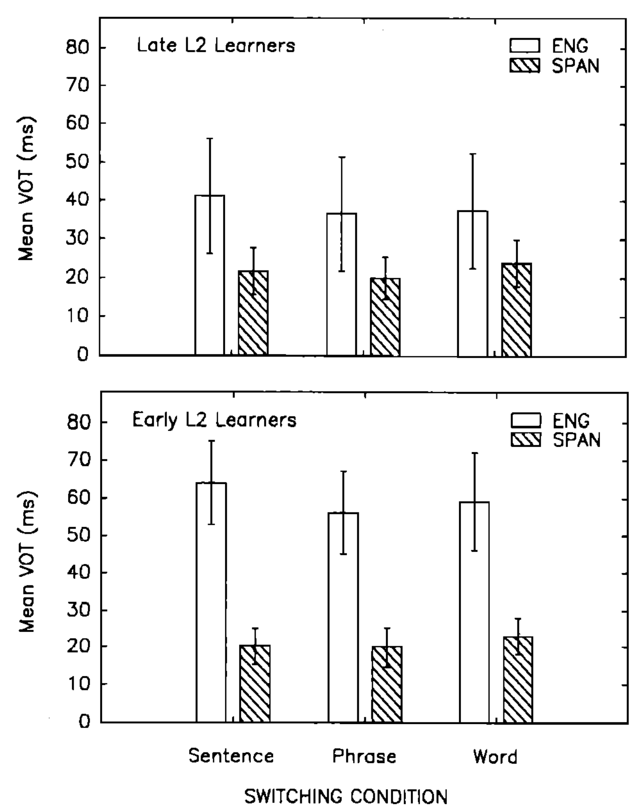

FIG. 3. The mean VOT (in ms) of Spanish and English /t/s spoken by early and late learners in three conditions (sentence, phrase, word) in which English and Spanish utterances were produced in alternation. Each mean is based on measures made for seven subjects in two vowel contexts. The error bars enclose +/− one standard deviation.

Averaged across the two groups and two vowel contexts, the magnitude of the Spanish vs English difference decreased as the rate of cross-language switching increased. It averaged 32 ms in the sentence condition, where the interval between successive Spanish and English /t/'s was the greatest. It averaged 26 ms in the phrase condition, and 25 ms in the word condition. These differences resulted in a significant language × switching condition interaction [$F(2,24) = 4.44$, $p < 0.05$], which was explored by tests of simple main effects. The condition factor was significant for both the English /t/ and the Spanish /t/ [$F(2,26) = 5.03$, 6.02, $p < 0.05$]. *Post hoc* tests revealed that VOT values were significantly greater for English /t/ in the sentence condition than in either the phrase or the word conditions (53 vs 46, 48 ms). VOT for Spanish /t/, on the other hand, was significantly *shorter* in the sentence and phrase conditions than in the isolated word condition (21, 20 vs 24 ms).

As in experiment 1, the VOT for /t/ was longer for stops preceding /i/ than /ɛ/. The vowel context factor was significant for both English /t/ and Spanish /t/ [$F(1,13) = 30.8$, 21.0; $p < 0.05$]. It is likely that the significant language × vowel context interaction obtained [$F(1,12) = 6.03$, $p < 0.05$] was due to the fact that, as in experiment 1, the vowel context effect was greater in English words (55 vs 44 ms) than in Spanish words (24 vs 19 ms).

2. Switching time

The early and late learners were given the same instructions, and told to time successive utterances in the three conditions in synchrony with light flashes emitted by a timer. The time between successive flashes decreased across conditions, which meant that the subjects had to switch ever more rapidly between languages. The lack of a significant group × language × condition interaction suggested that the need to switch more rapidly between L1 and L2 affected speech production by the two bilingual groups in the same way. However, it appears that the subjects in the two groups did not perform the task in the same way even though they were given the same instructions.

Although the subjects developed a rhythm based on the light flashes, the onsets of their utterances sometimes preceded the light flashes, and some utterances encroached onto the next interval. As an estimate of the time available for planning the production of each Spanish and English /t/, "/t/-to- /t/" intervals were measured to the nearest 0.1 ms in each condition from the release burst of the /t/ in one (Spanish or English) test word to the release burst of /t/ in the following test word. Two mean values were calculated for each subject in the three conditions. One mean value was based on measures of the time between the /t/ in each English test word and the /t/ in the following Spanish test word; the other mean value was based on the time between the /t/ in each Spanish test word and the /t/ in the following English test word. These intervals were designated the "E–S" and "S–E" switches, respectively. Mean E–S and S–E values were calculated for each subject for each of the three conditions. Each mean was based on two observations, one in which the vowel of the test word was /i/ and one in which the vowel was /ɛ/.

The measured duration of the intervals between /t/'s initiating successive Spanish and English test words was somewhat shorter on the average than the actual intervals specified by the light flashes in the sentence and phrase conditions (2.9 and 1.6), but slightly longer (1.2 s) in the word condition. The mean intervals for the early and the late learners are shown in Fig. 4 as a function of the direction of the switches. For the E–S switches, the /t/-to-/t/ intervals decreased as the rate of switching became more rapid (sentence: 2.7, phrase: 1.7, word: 1.2 s). The same held true for the S–E switches (sentence: 3.0, phrase: 1.5, word: 1.2 s). Averaged across the two groups, the /t/-to-/t/ intervals averaged 62 ms longer (i.e., 1.909 vs 1.847 s) for the S–E than E–S switches. The /t/-to-/t/ intervals averaged 312 ms longer (i.e., 2.034 vs 1.722 s) for the late than early learners. The difference between the two groups decreased from the sentence to the phrase to the word condition (497, 341, 100 ms).

The 84 mean /t/-to-/t/ values obtained for subjects in the two bilingual groups were submitted to a (2) group × (3) condition × (2) direction-of-switch ANOVA, with repeated measures on the last two factors. A significant group × condition interaction was obtained [$F(2,24) = 6.24$, $p < 0.05$], which was explored by testing the simple main effect of group for the three switching conditions. The early learners' /t/-to-/t/ intervals were significantly shorter than the late learners' in the sentence and phrase conditions [$F(1,12) = 10.18$, 9.82; $p \leqslant 0.05$] but not in the word condition [$F(1,12) = 2.66$].

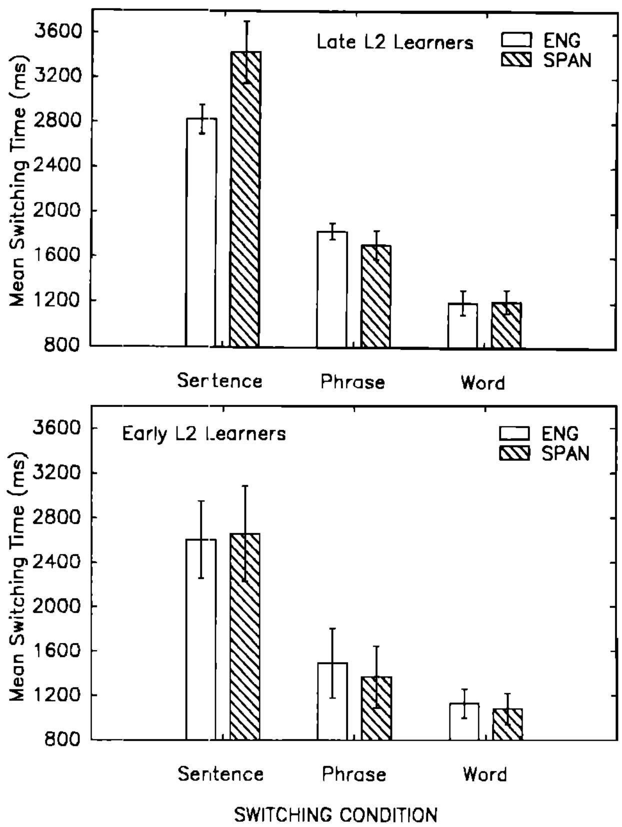

FIG. 4. The mean duration of the interval between the release bursts of successive Spanish and English /t/'s in three conditions (sentence, phrase, word) in which subjects were required to switch between Spanish and English. "SPAN" indicates switches from Spanish to English, "ENG" the English to Spanish switches. The error bars enclose + / − one standard deviation.

The direction-of-switch factor was marginally significant [$F(1,12) = 4.61, p = 0.053$]. A three-way interaction involving the direction factor was obtained [$F(2,24) = 10.4, p < 0.05$]. The simple main effect of group was tested for all eight direction × condition combinations. Only one between group difference was significant at a per-experiment error rate of 0.05 [$F(1,12) = 16.0, p = 0.0018$]. For S–E switches in the sentence condition, the /t/-to-/t/ intervals were significantly longer (by 772 ms) for the late than early learners. The two groups showed no significant differences when switching from English to Spanish.

C. Discussion

This experiment yielded results that were much the same as those obtained in experiment 1. Both early and late learners produced /t/ with significantly longer VOT values in English than Spanish, but the magnitude of the Spanish versus English VOT difference was substantially greater for the early than the late learners (41 vs 16 ms).

The bilingual subjects produced a larger VOT difference in the sentence condition, where Spanish and English sentences were produced in alternation, than in the word condition, where isolated Spanish and English words were produced in alternation. The differing size of the Spanish versus English difference occurred because the VOT of English /t/ decreased, whereas the VOT of Spanish /t/ increased somewhat, as the switching rate increased from the sentence to the word condition. Given the opposite direction of the changes in VOT for the Spanish /t/ and the English /t/, the changes were unlikely to have been caused simply by a change in speaking rate.

The group × condition × language interaction was nonsignificant. This might be taken to mean that the early and late learners switched between English and Spanish in the same way, but such a conclusion would be misleading. Even though VOT changed in much the same way for the two groups, and even though the same nominal procedures were followed for both groups, there was evidence that early and late learners did not perform the switching task in the same way. The late learners' /t/-to-/t/ intervals averaged 312 ms longer than the early learners'.

Previous research has shown that a measurable amount of time is needed to switch between two languages (Kolers, 1966). Macnamara *et al.* (1968) found that it took French–English bilinguals 210 ms longer to rapidly name lists of digits in French and English than to name digits in just one language. However, the duration of the /t/-to-/t/ intervals should probably not be regarded as a measure of the time needed to switch between languages, that is, to turn one language system "on" and the other "off" (Kolers, 1966). This is because, apart from the isolated words, the time from the onset of successive utterances and the (utterance-medial) /t/'s that were measured may have varied.[8]

The difference between the early and late learners' /t/-to-/t/ intervals implies that the late learners prolonged certain sounds, or paused to a greater extent, than the early learners. Some of the English sounds differed from any sound in the Spanish phonetic inventory. The English carrier sentence "Take another word such as _" contained a vowel and a consonant (viz. /eⁱ/,/t/) that only partially resemble any Spanish vowel or consonant. It contained three vowels and one consonant (viz. /æ//ɜ/,/ʌ/,/ð/) without a direct counterpart in Spanish; and it contained three consonants (/z/,/č/,/d/) that are not found in a comparable syllable position in Spanish.

It is tempting to speculate that the late learners may have prolonged sounds or paused between words in anticipation of the need to produce the non-Spanish sounds. They may have needed additional time to formulate plans for implementing English sounds that are not found in the Spanish phonetic inventory. The added time may have been expended on producing sounds whose mode of motoric implementation had not yet been fully automaticized, or to make modifications of previously established "programs" for implementing sounds in Spanish. One result obtained in experiment 2 is consistent with either interpretation. The /t/-to-/t/ intervals were 62 ms longer on the average when the bilingual subjects switched from English to Spanish than the reverse.

One possibility is that the "switching time" difference between early and late learners derived from the method used to elicit production in Spanish and English (viz. reading). However, Macnamara (1969) reported that, although reading speed is a strong predictor of relatively proficiency in two languages, the *speed* of switching between languages is not. Moreover, Macnamara *et al.* (1968) found no difference in switching time between two groups of bilinguals who apparently differed in much the same way as the early and late learners of the present study.

Finally, it is worth noting that experiment 2 examined code switching rather than code *borrowing*, defined by Grosjean and Soares (1986) as the production in a host language of a word/phrase from a donor language using the "phonology" of the host language. Their spectrographic data suggested, for example, that a native French speaker who inserts an English word into a French sentence will say the English word with French acoustic phonetic characteristics. The results presented here suggest that the early and late learners may have code switched in the same manner, at least in regards to speech production. The possibility exists, however, that the early and late learners would have been found to differ had their production of English words inserted into a Spanish conversation (or the reverse) been examined.

III. GENERAL DISCUSSION

The results obtained in experiments 1 and 2 confirmed a result obtained for late L2 learners in many previous studies of L2 speech production. Native Spanish speakers who had learned English as adults produced English /t/ with compromise VOT values intermediate to the values observed for Spanish and English monolinguals. The two experiments showed that native speakers of Spanish who learned English as young children, on the other hand, fully differentiated the /t/ of English and Spanish. Neither the early nor the late learners' VOT values differed from Spanish monolinguals' in the production of Spanish /t/. The early learners produced English /t/ with VOT values that were significantly longer—and therefore more English-like—than the late learners'. In fact, the early learners did not differ from English monolinguals in producing English /t/.

The present study is apparently the first to have compared the performance of early and late learners. The early learners were known to have received input from native English speakers when they first began to learn English. Two previous studies showed that early learners produced English /p,t,k/ with "compromise" VOT values (Caramazza *et al.*, 1973; Flege and Eefting, 1987a). The nature of the L2 phonetic input received by these early learners is not known. The results of the present study suggest that they may have received *accented* L2 input.

How can we account for the difference between the early and late learners in the production of English /t/? Three general types of explanation are possible. (1) The early learners may have perceived the acoustic phonetic characteristics of English voiceless stops more accurately than the late learners. This, in turn, might have prevented the late learners from producing English /t/ with native-like VOT values. (2) The early and late learners may have perceived English /t/ in the same manner, but the late learners may have been relatively less able than the early learners to motorically output what they represented perceptually. (3) The early and late learners may have been equally *able* to perceive English /t/ accurately and to develop means for producing voiceless stops with long-lag VOT values. However, the late learners may have been unable to fully *utilize* their sensorimotor capabilities owing to the state of development of their phonetic system when they began learning English as an L2. These three types of explanation will be considered in turn in the following sections.

A. Perception of L2 sounds

As discussed in the Introduction, Spanish and English /t/ differ in terms of glottal–supraglottal timing and place of linguapalatal constriction. Cross-language differences in speech production often coincide with differences in phonetic perception. Experiments with synthetic VOT continua, for example, have shown that longer VOT values are needed for native English than Spanish listeners to judge stops as voiceless (Abramson and Lisker, 1973; Williams, 1977a; Flege and Eefting, 1986). This perceptual difference corresponds to the longer VOT values observed in the production of English than Spanish /p,t,k/. Similarly, Elman *et al.* (1977) showed that Spanish monolinguals judged natural short-lag stops as /p/, whereas native English monolinguals judged them as /b/.

Bohn and Flege (1990), on the other hand, found that native speakers of English often judged short-lag Spanish [t] tokens as /t/. Native speakers of Spanish—even those with little previous exposure to English—consistently classified long-lag realizations of English /t/ as voiceless (see also Munro, 1987; Yeni-Komshian *et al.*, 1968). These findings suggest that VOT may be a less important cue to word-initial stop voicing contrasts than is commonly supposed.[9] More importantly for the current discussion, it suggests that Spanish and English /p,t,k/ are regarded as *phonologically* the same despite differences in VOT. Although Spanish and English /t/ differ acoustically, they seem to share certain properties, such as a lack of voicing immediately following stop release (Williams, 1977a), which causes bilinguals to identify them with one another.

An important issue for L2 research is whether the age at which L2 learning commences will affect how much acoustic phonetic information in L2 sounds is filtered out. Burnham (1986) suggested that certain phonemically nonrelevant acoustic dimensions are more easily perceived by listeners of all ages because they are salient auditorily (see also Best *et al.*, 1988). The acoustic phonetic contrast between /p,t,k/ in Spanish and English might be auditorily salient for Spanish learners of English, so that the acoustic phonetic differences between Spanish and English /t/ can be detected readily. The results obtained in a foreign accent mimicry experiment by Flege and Hammond (1982) suggested that native speakers of English can detect acoustic differences between the English /t/'s produced by native and Spanish speakers of English (see also Flege, 1984). VOT values were significantly longer in /t/'s spoken in normal English utterances than in utterances produced with a mimicked Spanish accent.

The differences between phonetic and phonemic processing suggests that the conscious perception of sound-sized units occurs primarily at the end of several processing stages. "Within-category" acoustic phonetic differences between English /p,t,k/ and Spanish /p,t,k/ may normally go unnoticed at a conscious level during the on-line comprehension of spoken language, but listeners may be able to gain access to phonetic information, or to exploit it in certain auditory processing tasks. That is, even though Spanish speakers may regard English /p,t,k/ as the "same" as Spanish /p,t,k/ at a phonemic level, they may treat the realizations of these phonemes as different at a phonetic or an auditory level of processing.

Many late learners examined in the present study approximated the English phonetic norm for /t/ without actually achieving it. The basis for the apparent limitation on how closely L2 phonetic norms were approximated may have been perceptual in nature. According to Flege (1988a, 1990a; Flege and Eefting, 1987a), "similar" sounds such as the /t/ of Spanish and English will be equated at a phonetic category level if L2 learning begins after about the age of 5 or 6 years, so that a distinct perceptual representation for English /t/ will not be developed. A number of "language set" experiments are consistent with the view that late learners of English L2 do not establish distinct perceptual representations for English voiceless stops. Elman *et al.* (1977) found that Spanish–English bilinguals who pronounced English with a foreign accent did not label short-lag stop differently (i.e., as /p/ vs /b/) when they were processing the stops in Spanish and English perceptual sets. Flege and Eefting (1987b) examined Dutch subjects' identification of the members of a VOT continuum ranging from /d/–/t/. Phoneme boundaries obtained in Dutch and English language sets differed significantly, but the size of the phoneme boundary shifts were much smaller (3 ms) than the difference one would expect between Dutch and English monolinguals.

There is some indirect evidence, on the other hand, that early learners may establish a perceptual representation for English /t/. The identification functions obtained for early learners by Caramazza *et al.* (1973) were nonmonotonic, showing shifts at two points along the VOT continua examined. A partial shift occurred at a point along the continua where French phoneme boundaries would be expected; complete shifts were observed at points nearer to the English phoneme boundary. Some subjects in the Elman *et al.* study showed a language set effect. These were the subjects who spoke English with little accent, so they may have been early learners (Diehl, 1988).

Other evidence suggests, however, that early learners do *not* establish a separate perceptual representation for English /t/, or do so at the expense of their previously established perceptual representation for Spanish /t/. The mean phoneme boundaries obtained for early learners in a forced-choice test by Caramazza *et al.* (1973) occurred at values intermediate to those observed for French and English monolinguals. Some of the Spanish–English bilinguals examined by Williams (1977b) had phoneme boundaries near those of Spanish monolinguals; others had boundaries nearer to those of English monolinguals.

Two experiments using synthetic stimuli with early learners failed to show significant language set effects with early learners (Caramazza *et al.*, 1974; Williams, 1977b). Bohn and Flege (1990) found the same, small language set effect for naturally produced short-lag stops for both early and late learners. Williams' (1979, 1980) research with Spanish–English bilingual children suggested that although they may initially divide VOT continua like Spanish monolinguals, their phoneme boundaries will shift toward English values as they gain experience in English. Also, sensitivity to the (Spanish) distinction between stops with lead and short-lag VOT in discrimination tests seemed to diminish, especially that of the children who learned English in early childhood.

The results obtained in previous speech perception research with early and late learners suggest two hypotheses concerning the perception of similar sounds in L1 and L2. The first is that neither early nor late learners establish perceptual representations for English /p,t,k/ that are distinct from representations established previously for the L1 /p,t,k/. On this view, the existing L1 categories either go unmodified, or else evolve so as to reflect the acoustic properties of voiceless stops in both L1 and L2.

The second hypothesis is that the results obtained in previous experiments employing a two-alternative forced-choice task, especially those employing synthetic stimuli, do not provide insight into how stops are specified at a *phonetic category* level. The existence of phonetic category representations may be difficult to demonstrate in a task that encourages listeners to access the final stage of auditory perceptual processing, that is, the phonemic stage. It has been suggested that the acoustic phonetic contrast between /t/ and /d/ in English is more salient auditorily than the Spanish phonetic contrast between /t/ and /d/ (Williams, 1980). If so, then if bilingual subjects *did* have distinct perceptual representations for short-lag and long-lag voiceless stops, they might have preferred to use the one that underlies the English /t/–/d/ distinction in a two-alternative forced-choice test.

The results obtained to date do not make it possible to choose between these two hypotheses. However, the finding of the present study that early but not late learners fully differentiated the /t/'s of Spanish and English is more consistent with the hypothesis that early learners *do* establish a perceptual phonetic representation for English /t/.

B. Speech learning ability

Even if one assumed that the late learners examined in the present study did have perceptual representations for English /t/, and those representations were as accurate as those of the early learners, the late learners might have differed from native speakers of English because they had passed a critical period for learning new forms of pronunciation. Some have supposed (e.g., Sapon, 1952) that pronunciation ability declines with age. Lenneberg (1967) concluded that a foreign accent in an L2 is "inevitable" if it is learned after puberty because brain development and lateralization for language function have reached completion by that time.[10] Many others have also hypothesized the existence of a critical period for human speech learning that derives from brain maturation (Penfield and Roberts, 1959; Lamendella, 1977; Scovel, 1988; but see Flege, 1987b).

One might hypothesize, therefore, that the late learners had less ability than the early learners to motorically implement their perceptual representations for sounds. There is, however, no direct evidence for age-related atrophy or change in those centers of the human brain that direct speech movements or regulate auditory processing. Moreover, the neural maturation hypothesis can be questioned on both neurological and empirical grounds (see Snow, 1987; Kinsbourne, 1981; Whitaker *et al.*, 1981; Krashen, 1973). Even if this were not so, an important problem exists for a critical period account.

A critical period account provides no insight into *which specific aspects* of the phonetic learning process may change with age. One wonders, for example, why the late learners showed compromise VOT values. There is no *a priori* reason to think that it is somehow easier for late learners to produce a partial modification of previously established articulatory patterns than to produce a *complete* modification that would enable them to match native speakers of English. In fact, the comparative rarity of stops with VOT values in the "compromise" range observed here for late L2 learners is probably disfavored for articulatory or perceptual reasons. Languages tend to have either short-lag stops or aspirated stops like those of English. Few languages, it seems, have stops with VOT values in between these two "modal" categories (Lisker and Abramson, 1964). But this is just what was observed here for late L2 learners.

C. Phonetic system depelopment

The speech learning model (SLM) described by Flege (e.g., 1988a, 1990a) can be used to account for why the early and late learners differed. As alluded to earlier, the SLM posits that auditory processing occurs at distinct auditory, phonetic, and phonemic levels (see also Werker and Logan, 1985; Burnham, 1986). The auditory level makes no reference to meaning or phonetic function. At the phonetic level, classes of phones are contrasted acoustically by properties that are sufficiently robust that they might be used to signal meaning contrasts in some language, but not necessarily the listener's native language (e.g., the difference between released and unreleased word-final stops in English). At the most abstract, phonemic, level of processing, phonetically relevant classes of phones ("sounds") are grouped together in functional units through the rules that comprise a learned, phonological system.

When processed at a phonemic level, sounds that may be distinct auditorily are treated as realizations of a single category. Fodor (1983) proposed that an important characteristic of input systems that make use of highly specialized, encapsulated systems is that the perceiver will have conscious awareness only of representations that are derived in the final stages of processing. However, according to the SLM, listeners remain able to access a phonetic level of representation, which enables them to learn to distinguish novel phonetic contrasts. If humans did not possess such an ability it would be impossible, for example, for students to learn to transcribe foreign sounds in phonetic classes or learn to note differences between allophones of a phoneme in their L1.

The SLM posits that speech production is organized at phonemic category, phonetic category, and sensory-motor levels. The phonemic categories specified in lexical entries are output using a finite number of universal phonetic categories that, in turn, are motorically output using phonetic realization rules. The model builds on work by Keating (1984), who described how phonetic and phonemic categories might be interfaced. Keating concluded that phonologically voiced and voiceless stops are implemented by one of three universal phonetic categories, corresponding roughly to Lisker and Abramson's three modal VOT categories. So, for example, the voiced phoneme /d/ may be implemented using short-lag or lead categories, and the voiceless phoneme /t/ may be implemented using a short-lag or a long-lag phonetic category.

One way in which English differs phonologically from Spanish in that the long-lag phonetic category is used to implement /t/ rather than the short-lag phonetic category. Language-specific realization rules are used to motorically output phonetic categories (Lieberman, 1970). The rules of one language, when applied to a long-lag stop category, might result in VOT that were slightly, but significantly longer, than the VOT values of another language.

The SLM posits that after phonetic categories have been established for L1 sounds in early childhood, listeners are increasingly likely to identify L2 sounds that partially resemble corresponding sounds in the L1 (referred to as "similar" sounds) as being realizations of an L1 category. Late learners will persist in identifying similar L2 sounds such as Spanish and English /t/, whereas early learners will eventually note the acoustic phonetic differences between them. As a result, early but not late learners will establish phonetic categories for similar L2 sounds, and early but not late learners will produce them authentically (i.e., like native speakers).

On this account, the early learners in the present study succeeded in fully differentiating Spanish /t/ and English /t/ because they used two different phonetic categories to implement the phoneme /t/ in Spanish and English. The late learners were unable to fully differentiate the /t/'s of Spanish and English because they did *not* have a separate phonetic category with which to implement /t/ in English. If the late learners did not possess a phonetic category for English /t/, how then did most of them produce it with longer VOT values than Spanish /t/?

The SLM posits that, when late learners identify corresponding L1 and L2 sounds in terms of a single phonetic category, but auditorily detect acoustic differences between them, they may produce the L1 and L2 sounds differently by applying different phonetic *realization rules.* According to Port and Mitleb (1983), realization rules determine the "details of speech timing and coordinate the commands to the speech articulators" (p. 220). The notion of "realization rule" is well established in the literature, but far more attention has been paid to temporal than spatial aspects of the gestures used to form speech sounds. This may be due to the fact that speech timing is often measured more easily than spatial properties such as the place of tongue–palate contact in stop consonants.

Most investigators have considered realization rules to be language specific (see, e.g., Liberman, 1970; Nooteboom, 1973; Klatt, 1976; Kent and Minifie, 1977; but cf. Stevens and House, 1972) perhaps owing to the small but systematic timing differences that have been observed between corresponding sounds in different languages (e.g., Lehiste, 1970; Ladefoged, 1980; Port *et al.*, 1980). Within a single language, phonetic realization rules are needed to account for how speakers systematically modify their production of a phonetic category as a function of, for example, social context (e.g., Labov, 1981). It is uncertain whether realization rules are distinct from the parameter manipulations used to effect changes in speaking rate or emphasis.

In fact, relatively little is known concerning the neural control mechanisms for phonetic realization rules. Lofqvist and his colleagues have examined in detail the production of stop consonants (Lofqvist, 1980; Lofqvist and Yoshioka, 1980, 1981; Lofqvist, 1980; Yoshioka *et al.*, 1981). This body of work indicates that the stereotypic laryngeal gesture used to ensure an interval of voicelessness in /p,t,k/ is effected by the coordinated innervation of intrinsic laryngeal muscles that rapidly abduct, then adduct, the vocal folds. In agreement with the earlier VOT research (Lisker and Abramson, 1964, 1967; Abramson, 1977), Lofqvist concluded that it is principally the *timing* of the laryngeal devoicing gesture with respect to supraglottal gestures which gives rise to a rich complex of acoustic features that includes VOT.

It is uncertain if the size of the glottal aperture or the rate of opening–closing can be regulated volitionally by talkers. It is also uncertain whether the timing of the devoicing gestures can be so regulated. Shaiman *et al.* (1985) found that, when the lip closing gestures for /p/ were delayed by an unanticipated perturbation, the laryngeal devoicing gestures were delayed proportionally. This suggests that the temporal coordination needed to specify language-specific VOT values may derive from rapid sensorimotor reflex interactions between laryngeal muscles such as the PCA and muscles used in forming supraglottal constrictions. The authors concluded that, although patterns of laryngeal–supralaryngeal coordination (and, ultimately, VOT) may represent a predetermined phonetic "goal," the timing pattern itself may not be "explicitly programmed but implemented downstream, by sensorimotor actions" (p. 185).

D. Critique of the three-level model

Speech production has historically been viewed as a two-stage process in which abstract sound units (phonemes) are first selected at higher levels of a message-generating system, then related to a lower level system for transformation into a code suitable for the generation and control of articulatory movement (e.g., Perkell, 1980; MacNeilage, 1980; MacNeilage *et al.*, 1981; Kent and McNeil, 1987; but see Lieberman, 1970). The speech production model proposed to account for differences between early and late learners differs from previous models in that it proposes *three* distinct levels of organization (Flege, 1988a, 1990a). That is, it distinguishes between a level of representation at which "universal" characteristics of phonetic segments are specified and a level at which fine-grained, language-specific detail is provided. The former is designated the level of *phonetic implementation*, the latter as the level of *phonetic realization*.

At the implementation level, a phoneme such as English /t/ would be represented as having tongue–tip constriction and a pattern of laryngeal timing that results in long-lag VOT values. The realization level would specify, among other things, a constriction of the tongue tip and dorsum against the alveolar ridge and a laryngeal timing pattern that results in VOT of approximately 65 ms rather than, say, 45 ms. This approach is consistent with the belief in the existence of universal sound types that are modified through language-specific learning (Chomsky and Halle, 1968).

One problem for this approach is that it rests on an untested assumption. With respect to the data of the present study, it is assumed that large VOT differences between Spanish [t] and English [t^h] phones can only arise through the implementation of /t/ using different phonetic categories. It is further assumed that relatively small VOT differences, such as the difference observed in the late learners' production of Spanish and English stops, will arise through the application of different realization rules to a single phonetic category.

Experiment 2 afforded the opportunity for testing the distinction between phonetic implementation and realization. The need to choose between competing structures adds a finite amount of processing time in motor tasks (e.g., Sternberg, 1969). If one assumes that lexical items are marked for language identity (Macnamara, 1969; Neufeld, 1976), and that phonetic implementation processes are readied when a word is lexically accessed for production (Flege, 1990c), then one might expect late learners to take slightly longer to motorically output L2 words containing similar sounds than early learners. This is because late but not the early learners would need to choose between two completing realization rules to motorically output Spanish and English words with /t/.

In experiment 2, the bilingual subjects were required to switch with increasing rapidity between Spanish and English. Under sufficient time pressure, the late learners might have been expected to abandon the later-acquired (English) realization rule in favor of the early-acquired (Spanish) realization rule (see Kewley-Port and Preston, 1974). Experiment 2 did not, however, provide support for the distinction between phonetic implementation and realization. The magnitude of the Spanish versus English VOT difference did decrease slightly as the rate of switching increased, but the size of the decrease was not significantly greater for the late learners than for the early learners.

This finding did not provide counterevidence against the implementation–realization distinction, however. The interval between successive English and Spanish /t/'s was over 300 ms longer for the late than the early learners. It is conceivable that this extra time was due to the late learners' need to select between competing realization rules. Additional research is needed to test this hypothesis, for it is also possible that the added time needed by the late learners was due to a need for accessing relatively unfamiliar L2 phonetic *categories*.

Finally, the present study did not provide direct evidence that the late learners *did not have* distinct categories for the /t/'s of Spanish and English. As noted earlier, one might argue that most of the late learners produced English /t/ with only slightly longer VOT values in English than Spanish—rather than the substantially longer VOT values seen for the early learners—because their phonetic categories for English /t/ were inaccurate. Or, one might argue that they were *less skillful* than early learners in implementing English /t/ as a long-lag stop.

E. Summary and conclusions

The present study provided evidence that Spanish–English bilinguals can fully differentiate Spanish and English /t/, at least in terms of VOT, if they learn English as an L2 in early childhood but not if they begin learning English as adults. A review of the literature provided no support for the view that late learners are inherently less capable of learning new forms of pronunciation. The literature review suggested that early learners may be more apt than late learners to develop a central perceptual representation for the long-lag stops used to implement English /t/, although this conclusion is by no means certain and no perception data were provided for the subjects in the present study.

The VOT difference between early and late learners for English /t/ was interpreted to reflect a difference in phonetic organization. Specifically, it was claimed that the early learners fully differentiated English /t/ from Spanish /t/ because they, unlike the late learners, had distinct phonetic categories for the two /t/'s. The late learners were hypothesized to produce their relatively small Spanish versus English VOT difference by using two different phonetic realization rules to output a single phonetic category. This interpretation of the data presented here must be considered tentative. No direct evidence was provided for the distinction drawn between phonetic implementation and realization. Moreover, although it appeared that the early learners were better able to prevent the L1 phonetic system from influencing their L2 speech production, factors that may influence the degree of independence of L1 and L2 systems are poorly understood and therefore may not have been controlled for adequately in the present study (Obler and Albert, 1978; Grosjean, 1985).

However, if the interpretation offered here is supported by additional research, it would suggest two important conclusions about the nature of bilingualism: (1) Early L2 learners may have an enriched phonetic system that includes all phonetic categories possessed by native speakers of the L1 and L2; and (2) a bilingual's two languages are represented by a unique system that does not represent the sum of the competences of two monolingual speakers. This second conclusion is based on the inference that the late learners developed a phonetic realization rule that neither Spanish monolinguals nor English monolinguals possess, namely, one used to slightly increase VOT in English as opposed to Spanish /t/.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant DC00257. Thanks are extended to R. Diehl and P. MacNeilage for help locating subjects in Austin; L. Cueva and C. Mena for help gathering data; G. Allen, M. Munro, and two anonymous reviewers for comments on a previous draft of this article; and L. Skelton for help with all phases of the research.

[1]Suomi (1980) found that Finnish high school students often produced English /p,t,k/ with Finnish-like short-lag VOT values. Schmidt (1988) examined the English spoken by three native Spanish L2 learners with a demonstrated "superior" pronunciation of English who had begun learning English between the ages of 10–13 years. These subjects produced English /p/ with shorter VOT values than a group of native English speakers (33, 40, and 51 vs 72 ms). Flege and Eefting (1987b) found that most Dutch adults who began learning English at the age of 12 years tended to produce English /t/ with short-lag VOT values if they had little English-language experience. Those who were experienced tended to produce English /t/ with compromise VOT values. Williams (1979, 1980) found that even native Spanish children who began learning English between the ages of 8–10 years produced English /p/ with shorter VOT values than native English children.

[2]Mack (1990) observed "overshooting" of the English phonetic norm by a 10 year old who had learned both French and English in early childhood. The child produced English /p,t,k/ with longer VOT values than native English children of the same age. He produced English /p,t,k/ with longer VOT values than French /p,t,k/ which, in turn, had longer VOT values than French /b,d,g/, which were unexpectedly produced with English-like short-lag VOT values. It appears that the child was attempting to differentiate L1 and L2 stops that differed phonetically but not phonologically.

[3]Arsenian (1937; cited by Macnamara, 1969) estimated that the reliability of subjects' self-reports concerning language usage was of the order of $r = 0.800$ or better.

[4]The Spanish /e/ phoneme is symbolized here as /ɛ/ because it is usually realized with an [ɛ] variant in closed syllables (Dalbor, 1980). The vowel in *tema* may actually have been an [e]-quality vowel because it occurred in an open syllable.

[5]The research assistant who elicited data from the native English and Spanish subjects in Austin was a native speaker of Spanish who had learned English at about the age of 5 years. She seemed to the author to speak English without an accent, and reported that this was also true of her Spanish. The research assistant who elicited data from six late learners in Birmingham, Alabama was a Honduran who had begun learning English at about the age of 12 years in a bilingual school. She spoke English with a slight accent in the author's opinion. Both assistants switched easily and naturally between the Spanish and English portions of the experiment, as required by the protocol.

[6]The standard deviations associated with utterance-medial stops were slightly greater than those associated with utterance-initial stops (10.1 vs 7.9), perhaps because the utterance-medial stops occurred in eight different words whereas the utterance-initial stops occurred in a single word (viz. *take*). The three-way interaction was probably significant because the vowel×utterance position interaction was significant for the early learners [$F(1,9) = 15.1, p < 0.05$] but not for the English monolinguals or the late learners. Further analysis revealed that, for the early learners, s.d. values were greater for utterance-medial /t/ tokens which preceded /i/ than /ɛ/ (13.5 vs 7.8). It came as no surprise that the vowel context effect was nonsignificant for utterance-initial /t/ (9.3 vs 8.6), for these stops were more distant from the differential vowel context in the test words at the end of the carrier phrase.

[7]There was no vowel context effect for the two sets of utterance-initial stops that were measured. The VOT of the utterance-initial Spanish stops in *tengo* had the same mean VOT value (19 ms) in utterances ending in words with /i/ and /ɛ/. This was expected because the differential vowel context was too distant to exert an effect on the utterance-initial stops. Similarly, the English stops in *take* had about the same mean VOT values in the utterances ending in test words with /i/ and /ɛ/ (45 vs 47 ms).

[8]A more direct measure of actual language-to-language "switching time" might have been a measure of utterance onset to utterance onset. This interval was not measured, however, owing to the problem of how to interpret inter- and intra-utterance pauses.

[9]Other cues such as burst and aspiration intensity, and F_1 onset frequency and transition duration, are known to play a role in the perception of stop voicing (see, e.g., Stevens and Klatt, 1974; Lisket *et al.*, 1977). Forrest and Rockman (1988) showed recently that English adults may correctly identify at least some of misarticulating children's short-lag realizations of English /p,t,k/ as voiceless. Conversely, some children perceived to neutralize the distinction between initial voiced and voiceless stops may nevertheless produce a significant VOT contrast between them (see, e.g., Catts and Jensen, 1983).

[10]There is nothing inherently implausible with a critical period hypothesis. For example, Nottebohm (1989) showed that the seasonal pattern of song learning, loss, and relearning in the male canary is related to changes in the size of two neural song control centers in the canaries' forebrain (see also Nottebohm, 1969).

Abramson, A. (1977). "Laryngeal timing in consonant distinctions," Phonetica **34**, 295–303.

Abramson, A., and Lisker, L. (1973). "Voice timing perception in Spanish word-initial stops," J. Phon. **1**, 1–8.

Altenberg, E., and Cairns, H. (1983). "The effects of phonotactic constraints on lexical processing in bilingual and monolingual subjects," J. Verb. Learn. Verb. Behav. **22**, 174–188.

Arsenian, S. (1937). "Bilingualism and Mental Development: A study of the intelligence and the social background of bilingual children in New York City," Teachers College, Columbia Univ.

Asher, J., and Garcia, R. (1969). "The optimal age to learn a foreign language," Mod. Lang. J. **53**, 334–341.

Best, C., McRoberts, G., and Sithole, N. (1988). "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," JEP:HPP **14**, 345–360.

Blair, D., and Harris, R. (1981). "A test of interlingual interaction in comprehension by bilinguals," J. Psycholing. Res. **10**, 457–467.

Bohn, O.-S., and Flege, J. (1990). "Perceptual switching in Spanish/English bilinguals: Evidence for universal factors in voicing judgments," J. Phon. (submitted).

Burnham, D. (1986). "Developmental loss of speech perception: Exposure to and experience with a first language," Appl. Psycholing. **7**, 207–240.

Caramazza, A., Yeni-Komshian, G., Zurif, E., and Carbone, E. (1973). "The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals," J. Acoust. Soc. Am. **54**, 421–428.

Caramazza, A., Yeni-Komshian, G., and Zurif, E. (1974). "Bilingual switching: The phonological level," Can. J. Psychol. **28**, 310–318.

Catts, H., and Jensen, P. (1983). "Speech timing of phonologically disordered children: Voicing contrast of initial and final stop consonants," J. Speech Hear. Res. **26**, 501–510.

Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English* (Harper & Row, New York)

Dalbor, J. (1980). *Spanish Pronunciation* (Holt, Rinehart, and Winston, New York).

Diehl, R. (1988). Personal communication.
Elman, J., Diehl, R., and Buchwald, S. (1977)."Perceptual switching in bilinguals," J. Acoust. Soc. Am. **62**, 971–974.
Ferguson, C. (1986). "Discovering sound units and constructing sound systems: It's child play," in *Invariance and Variability in Speech Processes*, edited by J. Perkell and D. Klatt (Earlbaum, Hillsdale, NJ), pp. 36–57.
Fernandez, M., and Molinet–Molina, M. (1988). "El español de los niños pequenos en Miami," Paper presented at the 9th Int. Conf. on Spanish in the U.S., Miami, October 1988.
Flege, J. (1984). "The detection of French accent by American listeners," J. Acoust. Soc. Am. **76**, 692–707.
Flege, J. (1987a). "The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification," J. Phon. **15**, 47–65.
Flege, J. (1987b). "A critical period for learning to pronounce foreign languages?," Appl. Ling. **8**, 162–177.
Flege, J. (1988a). "The production and perception of speech sounds in a foreign language," in *Human Communication and Its Disorders, A Review 1988*, edited by H. Winitz (Ablex, Norwood, NJ), pp. 224–401.
Flege, J. (1988b). "Factors affecting degree of perceived foreign accent in English sentences," J. Acoust. Soc. Am. **84**, 70–79.
Flege, J. (1990a). "The Intelligibility of English vowels spoken by British and Dutch talkers," in *Intelligibility in Speech Disorders: Theory, Measurement, and Management*, edited by R. Kent (Benjamins, Amsterdam) (in press).
Flege, J. (1990b). "At what age of learning is a foreign accent first evident?," submitted to J. Acoust. Soc. Am.
Flege, J. (1990c). "The production of cognate English words by native speakers of Spanish: More evidence for the distinction between phonetic implementation and realization," submitted to J. Phon.
Flege, J., and Eefting, W. (1986). "Linguistic and developmental effects on the production and perception of stop consonants," Phonetica **43**, 155–171.
Flege, J., and Eefting, W. (1987a). "The production and perception of English stops by Spanish speakers of English," J. Phon. **15**, 67–83.
Flege, J., and Eefting, W. (1987b) "Cross-language switching in stop consonant production and perception by Dutch speakers of English," Speech Commun. **6**, 185–202.
Flege, J., and Eefting, W. (1988). "Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation," J. Acoust. Soc. Am. **83**, 729–740.
Flege, J., and Hammond, R. (1982). "Mimicry of non-distinctive phonetic differences between language varieties," Stud. Sec. Lang. Acquis. **5**, 1–18.
Flege, J., and Hillenbrand, J. (1984). "Limits on pronunciation accuracy in adult foreign language speech production," J. Acoust. Soc. Am. 76, 708–721.
Flege, J., and Port, R. (1981). "Cross-language phonetic interference: Arabic to English," Lang. Speech **24**, 125–146.
Fodor, J. (1983). *The Modularity of Mind* (MIT, Cambridge, MA).
Fokes, J., Bond, Z., and Steinberg, M. (1985). "Acquisition of the English voicing contrast by Arab children," Lang. Speech **28**, 81–92.
Forrest, K., and Rockman, B. (1988). "Acoustic and perceptual analysis of word-initial stop consonants in phonologically disordered children," J. Speech Hear. Res. **31**, 449–459.
Grosjean, F. (1982). *Life with Two Languages, An Introduction to Bilingualism* (Cambridge U. P., Cambridge).
Grosjean, F. (1985). "The bilingual is a competent but specific speaker-hearer," J. Multiling. Multicultural Dev. **6**, 467–477.
Grosjean, F. (1989). "Neurolinguists, Beware! The bilingual is not two monolinguals in one person," Brain Lang. **36**, 3–15.
Grosjean, F., and Soares, C. (1986). "Processing mixed languages: Some preliminary findings," in *Language Processing in Bilinguals: Psycholinguistic and Neurophysiological Perspectives*, edited by J. Vaid (Erlbaum, Hillsdale, NJ).
Keating, P. (1984). "Phonetic and phonological representation of stop consonant voicing," Language **60**, 286–319.
Kent, R., and McNeil, M. (1987). "Relative timing of sentence repetition in apraxia of speech and conduction aphasia," in *Phonetic Approaches to Speech Production in Aphasia and Related Disorders*, edited by J. Ryalls (College Hill, San Diego, CA), pp. 181–220.
Kent, R., and Minifie, F. (1977). "Coarticulation in recent speech production models," J. Phon. **5**, 115–134.
Kewley-Port, D., and Preston, M. (1974). "Early apical stop production: A voice onset time analysis," J. Phon. **2**, 195–210.
Kinsbourne, M. (1981). "Neuropsychological aspects of bilingualism," in *Native Language and Foreign Language Acquisition*, edited by H. Winitz (Annals of the New York Academy of Sciences, New York), Vol. 379, pp. 50–58.
Klatt, D. (1976). "Linguistic uses of segmental duration in English: acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**, 1208–1221.
Kolers, P. (1965). "Bilingualism and bicodalism," Lang. Speech **8**, 122–126.
Kolers, P. (1966). "Reading and talking bilingually," Am. J. Psychol. **79**, 357–376.
Krashen, S. (1973). "Lateralization, language learning, and the critical period: Some new evidence," Lang. Learn. **23**, 63–74.
Labov, W. (1981). "Resolving the Neogrammarian controversy," Language **57**, 267–309.
Ladefoged, P. (1980). "What are linguistic sounds made of?," Language **56**, 485–502.
Lambert, W., and Rawlings, C. (1969). "Bilingual processing of mixed-language associative networks," J. Verb. Learn. Verb. Behav. **8**, 604–609.
Lamendella, J. (1977). "General principles of neurofunctional organization and their manifestation in primary and non-primary language acquisition," Lang. Learn. **27**, 155–196.
Lehiste, I. (1970). *Suprasegmentals* (MIT, Cambridge, MA).
Lenneberg, E. (1967). *Biological Foundations of Language* (Wiley, New York).
Liberman, A. (1970). "The grammars of speech and language," Cognit. Psychol. **1**, 301–323.
Lieberman, P. (1970). "Towards a unified phonetic theory," Ling. Inquiry **3**, 307–321.
Lisker, L., and Abramson, A. (1964). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–423.
Lisker, L., and Abramson, A. (1967). "Some effects of context on voice onset time," Lang. Speech **10**, 1–28.
Lisker, L., Liberman, A., Erickson, D., Dechovitz, D., and Mandler, R. (1977). "On pushing the voice-onset-time (VOT) boundary about," Lang. Speech **20**, 209–216.
Lofqvist, A. (1980). "Interarticulator programming in stop production," J. Phon. **8**, 475–490.
Lofqvist, A., and Yoshioka, H. (1980). "Laryngeal activity Swedish obstruent clusters," J. Acoust. Soc. Am. **68**, 792–801.
Lofqvist, A., and Yoshioka, H. (1981). "Interarticulator programming in obstruent production," Phonetica **38**, 21–34.
Lowie, W. (1988). "Age and foreign language pronunciation in the classroom," unpublished thesis, University of Amsterdam.
Mack, M. (1986). "A study of semantic and syntactic processing in monolinguals and fluent early bilinguals," J. Psycholing. Res. **15**, 463–488.
Mack, M. (1989). "Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals," Percept. Psychophys. **46**, 187–200.
Mack, M. (1990). "Phonetic transfer in a French-English bilingual child," in *Languages Attitudes and Language Conflict*, edited by P. Nelde (Dummler, Bonn), pp. 107–124.
Macnamara, J. (1969). "How can one measure the extent of a person's bilingual proficiency?," in *Description and Measurement of Bilingualism*, edited by L. Kelley (University of Toronto, Toronto, Canada), pp. 80–98.
Macnamara, J., Krauthammer, M., and Bolgar, M. (1968). "Language switching in bilinguals as a function of stimulus and response uncertainty," J. Exp. Psychol. **78**, 208–215.
MacNeilage, P. (1980). "Speech production," Lang. Speech **23**, 3–23.
MacNeilage, P., Hutchinson, J., and Lasater, S. (1981). "The production of speech: Development and dissolution of motoric and premotoric processes," in *Attention and Performance, IX*, edited by J. Long and A. Baddeley (Earlbaum, Hillsdale, NJ), pp. 503–519.
Magiste, E. (1979). "The competing language systems of the multilingual: A developmental study of decoding and encoding processes," J. Verb. Learn. Verb. Behav. **18**, 79–89.
Major, R. (1987). "English voiceless stop production by speakers of Brazilian Portuguese," J. Phon. **15**, 197–202.
Morosan, D., and Jamieson, D. (1989). "Evaluation of a technique for training new speech contrasts: Generalization across voices, but not word-position or task," J. Speech Hear. Res. **32**, 501–511.
Munro, M. (1987). "Voicing Contrasts in French and English Bilabial Stops," unpubl. M.Sc. thesis, Univ. of Alberta, Edmonton, Alberta.
Nathan, G. (1987). "On second-language acquisition of voiced stops," J. Phon. **15**, 313–322.

Neufeld, G. (1976). "The bilingual's lexical store," Int. Rev. Appl. Linguist. **14**, 15–35.

Nooteboom, S. (1973). "Perceptual reality of prosodic durations," J. Phon. **1**, 25–45.

Nottebohm, F. (1969). "The 'critical period' for song learning," Ibis **111**, 386–387.

Nottebohm, F. (1989). "From bird song to neurogenesis," Sci. Am., February 1989, 74–79.

Obler, L., and Albert, M. (1978). "A monitor system for bilingual language processing," in *Aspects of Bilingualism*, edited by M. Paradis (Hornbeam, Columbia, SC), pp. 156–164.

Oyama, S. (1976). "The sensitive period for the acquisition of a nonnative phonological system," J. of Psycholing. Res. **5**, 261–285.

Penfield, W., and Roberts, L. (1959). *Speech and Brain Mechanisms* (Princeton U.P., Princeton, NJ).

Perkell, J. (1980). "Phonetic features and the physiology of speech production, " in *Language Production, Vol. 1*, edited by B. Butterworth (Academic, New York), pp. 337–372.

Port, R., Ál-ani, S., and Maeda, S. (1980). "Temporal compensation and universal phonetics," Phonetica **37**, 235–252.

Port, R., and Mitleb, F. (1980). "Phonetic and phonological manifestations of the voicing contrast in Arabic-accented English," Res. Phon. **1**, 137–165 (Dept. of Linguistics, Indian University).

Port, R., and Mitleb, F. (1983). "Segmental features and implementation in acquisition of English by Arabic speakers," J. Phon. **11**, 219–231.

Port, R., and Rotunno, R. (1979)."Relation between voice-onset time and vowel duration," J. Acoust. Soc. Am. **66**, 654–662.

Purcell, E., and Suter, R. (1980). "Predictors of pronunciation accuracy: A reexamination," Lang. Learn. **30**, 271–287.

Sapon, S. (1952). "An application of psychological theory to pronunciation problems in second language learning," Mod. Lang. J. **36**, 111–114.

Schmidt, A. (1988). "The acquisition of some American English duration parameters by nonnative speakers of English," unpubl. Ph.D. dissertation, Univ. of Florida.

Scovel, T. (1988). *A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech* (Harper & Row, New York).

Seliger, H., Krashen, S., and Ladefoged, P. (1975). "Maturational constraints in the acquisition of second language accent," Lang. Sci. **36**, 20–22.

Selinker, L. (1972). "Interlanguage," Int. Rev. Appl. Ling. **10**, 209–231.

Shaiman, S., Abbs, J., and Gracco, V. (1985). "Sensorimotor contributions to oral-laryngeal coordination for speech," Soc. for Neurosci. Abst. **11**, 76.

Snow, C. (1987). "Relevance of the notion of a critical period to language acquisition," in *Sensitive Periods in Development*, edited by M. Bornstein (Erlbaum, Hillsdale, NJ), pp. 183–210.

Sternberg, S. (1969). "Memory-scanning: Mental processes revealed by reaction time experiments," Am. Sci. **57**, 421–457.

Stevens, K., and Klatt, D. (1974). "The role of formant transitions in the voiced-voiceless distinction for stops," J. Acoust. Soc. Am. **55**, 653–659.

Stevens, K., and House, A. (1972). "Speech perception," in *Foundations of Modern Auditory Theory, Vol. 2*, edited by J. Tobias (Academic, New York).

Suomi, K. (1980). "Voicing in English and Finnish stops," Publ. 10 of the Dept. of English and General Linguistics, Turku Univ., Finland.

Tahta, S., Wood, M., and Lowenthal, K. (1981). "Foreign accents: Factors relating to transfer of accent from the first language to the second language," Lang. Speech **24**, 265–272.

Thompson, I. (1984). "Experimental study of foreign accents," unpubl. Ph.D. thesis, George Washington Univ., Washington, DC.

Trubetzkoy, N. (1939). *Grundzuge der Phonologie* (University of California, Berkeley); English translation by C. Baltaxe, 1969.

Weinreich, U. (1953). *Languages in Contact, Findings and Problems* (Mouton, The Hague).

Weismer, G. (1979). "Sensitivity of voice-onset time (VOT) measures to certain segmental features in speech production," J. Phon. **7**, 197–204.

Werker, J., and Logan, J. (1985). "Cross-language evidence for three factors in speech perception," Percept. Psychophys. **37**, 35–44.

Whitaker, H., Bub, D., and Leventer, S. (1981). "Neurolinguistic aspects of language acquisition and bilingualism," in *Native and Foreign Language Acquisition*, edited by H. Winitz (New York Academy of Sciences, New York), pp. 59-74.

Williams, L. (1977a). "The voicing contrast in Spanish," J. Phon. **5**, 169–184.

Williams, L. (1977b). "The perception of stop consonant voicing by Spanish-English bilinguals," Percept. Psychophys. **21**, 289–297.

Williams, L. (1979). "The modification of speech perception and production in second-language learning," Percept. Psychophys. **26**, 95–104.

Williams, L. (1980). "Phonetic variation as a function of second-language learning," in *Child Phonology, Vol. 2 Perception*, edited by G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (Academic, New York), pp. 185–216.

Yeni-Komshian, G., Preston, M., and Kewley-Port, D. (1968). "A cross-language study of the perception of stop consonants in American and Lebanese children," Neurocommun. Lab., Johns Hopkins Univ. **3**, 1–15.

Yoshioka, H., Lofqvist, A., and Hirose (1981). "Laryngeal adjustments in the production of consonant clusters and geminates in American English," J. Acoust. Soc. Am. **70**, 1615–1632.